An investigator collected data on heights and weights of college students; results can be summarized as follows.

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men’s height</td>
<td>70 inches</td>
<td>3 inches</td>
</tr>
<tr>
<td>Men’s weight</td>
<td>144 pounds</td>
<td>21 pounds</td>
</tr>
<tr>
<td>Women’s height</td>
<td>64 inches</td>
<td>3 inches</td>
</tr>
<tr>
<td>Women’s weight</td>
<td>120 pounds</td>
<td>21 pounds</td>
</tr>
</tbody>
</table>

The correlation coefficient between height and weight for the men was about 0.60; for the women it was about the same. If you take the men and women together, the correlation between height and weight would be __________.

just about 0.60  somewhat lower  somewhat higher

Choose one option and explain.

The correlation coefficient of the combined data will be higher than 0.6. The reasoning is that the combined data will follow a line more dramatically than the individual data sets. This can be seen visually in the following picture. The “new” point of averages is somewhere in the overlap of the two clouds as shown below.

The clouds lie primarily in the positive quadrants around this new point of averages, much more so than before when we were considering the data for men and women separately. Note that this is not a mathematical proof; indeed, this would be hard to show rigorously.
In the article titled “Educating women saves millions of kids” (link provided on Homework view) a negative correlation is found between the death rate for children and the average years of education for women in 175 developing countries from 1970 to 2009.

(a) Give one example of confounding variable that could explain the association observed between average years of education for women and death rates for children. Then clearly explain why its effect on death rate may be confounded with the effect of education. (1 pts)

One variable that may explain the association between education of women and death rates of children could be the economic development in a country. For the effect of this variable to be confounded with the effect of education, we would need to see that (a) the countries where the women are attaining higher levels of education are precisely the countries where economic development is in full swing, and (b) economic development leads to a better quality of life and improved health care and hence a reduction in death rates of children.

(b) Does the association found in the article provide a fair assessment of the association between education and death rates? Explain your answer. (1 pts)

No, the association does not provide a fair assessment of the association between education and death rates because it is not countries that get education and it is not countries that die. People need to be the observational unit for us to get a reliable measure of the impact of education on mortality. In particular, our observational unit should be the individual women in these countries. On each woman, we must measure the level of her education as well as the number of children she may have lost.

(c) Use what you learn from parts (a) and (b) to figure out why the title is misleading. Then compose a brief letter to the reporter of the article explaining why her title is misleading. (1 pts)

The title of your article is misleading as it gives the impression that the education of women causes a reduction in child mortality. The first problem with this conclusion is that the data in the study you describe is from an observational study and no attempts have been (or can be) made to eliminate the effects of other potentially relevant variables. For instance, we might be simply observing the effects of economic development in these countries rather than a cause-effect between education and mortality. Economic development results in both improved education facilities and improved health care. The second problem is that this study does not give us a fair assessment of the strength of the association between child mortality and the education of women. In order to study this association, the researchers would need to collect data on the individual women themselves, a daunting task. Nevertheless, it does not mean that this research is not important or illuminating. It simply means that your conclusions are too strong and not supported by the data at hand.

In a large psychology class, each subject took two I.Q. tests (form L and form M of the Stanford-Binet). A scatter diagram for the test scores is shown below. You are trying to predict the score on form M from form L. Each prediction is off by some amount. On the whole, will these prediction errors be smaller when the score on form L is 75, or 125? or is it about the same for both? Explain your answer. (1 pts)
On the whole, prediction errors will be smaller when the score on form L is 75 than when it is 125. For some reason, scores on form M seem not to vary much when the score on L is around 75. And since the prediction error is the typical size of the variation from our estimate, we can conclude that this error will be smaller when around 75.

4. Pearson and Lee obtained the following results in a study of about 1,000 families.

\[
\begin{align*}
\text{average height of husband} & \approx 68 \text{ inches, } \text{ SD } \approx 2.7 \text{ inches}, \\
\text{average height of wife} & \approx 63 \text{ inches, } \text{ SD } \approx 2.5 \text{ inches}, \\
r & \approx 0.25
\end{align*}
\]

The scatter plot is football shaped. Tomorrow a wife will be selected at random. Predict her height if the height of her husband is:

(a) 72 inches: The husband who measures 72 inches is 1.5 SDs above the average height of the husbands. So, by regression, his wife will be \(0.25 \times 1.5\) or 0.4 SDs above the average height of the wives. That is, she will be 64 inches on average. So I would predict 64 inches.

(b) 64 inches: The husband who measures 64 inches is 1.5 SDs below the average height of husbands. So by regression, his wife will only be \(0.25 \times 1.5\) or 0.4 SDs below the average height of the wives. That is, she will be 62 inches tall on average. So I would predict 62 inches.

(c) unknown: In this case, our prediction would be the average height of wives which is 63 inches. Since we have no information on her husband’s height, we do not know where in the cloud this data point will fall. In the absence of such information, our best bet is to go with the center or what is “typical” for the wives.

Show all work. (1.5 pts)
5. A doctor is in the habit of measuring blood pressure twice. She notices that patients who are unusually high on the first reading tend to have somewhat lower second readings. She concludes that patients must be more relaxed on the second reading. Explain in simple language the fallacy in her thinking. You may assume that the correlation between the two sets of blood pressure readings is positive but moderate, at best. (0.5 pts)

This is a case where nothing exciting may be going on and the data is simply reflecting the weak correlation between successive blood pressure readings. That is, by virtue of the small correlation seen with blood pressure readings, readings are not as extreme on the second measurement as they are on the first measurement. That is, there really is not a very strong linear relationship between successive measurements. So her approach of trying to find some meaning to how the second set of measurements relate to the first is making a mountain out of a molehill. For example, another doctor might focus on the patients who had low readings the first time and note that they suddenly ended up with higher readings on the second time. Does this mean that patients were relaxed and then became tense? Not really. There just isn’t a strong connection between successive measurements.